Exogenous Predictors of National Performance Measures for Emergency Department Crowding

Jesse M. Pines, MD, MBA, Sandra L. Decker, PhD, Tianyan Hu

From the Departments of Emergency Medicine and Health Policy, George Washington University, Washington, DC (Pines); the National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD (Decker); and the Department of Economics, Lehigh University, Bethlehem, PA (Hu).

Study objective: We explore the relationship between exogenous-level predictors and performance on 4 emergency department (ED) throughput measures approved by the National Quality Forum: median ED length of visit for admitted and discharged patients, median waiting time, and rate of left without being seen. We seek to find predictors for benchmarking and public reporting.

Methods: This was a study of 424 US hospitals that reported data to the National Hospital Ambulatory Care Survey in 2008 to 2009. Wald F tests and generalized linear models were used to test the relationship between exogenous variables (case mix, age mix, ED volume, teaching status, and Metropolitan Statistical Area status) and performance on the measures.

Results: Median waiting time was 35 minutes (95% confidence interval [CI] 26 to 43 minutes), median length of visit for patients treated but not admitted was 131 minutes (95% CI 121 to 142 minutes), median length of visit for patients admitted was 244 minutes (95% CI 218 to 270 minutes), and rate of left without being seen was 1.3% (95% CI 0.9% to 1.8%). Most exogenous variables, including ED volume, Metropolitan Statistical Area, teaching hospital status, age mix, and case mix, demonstrated significant association with waiting times and lengths of visit. Older age and a higher proportion of respiratory complaints were associated with differences in rates of left without being seen.

Conclusion: Several exogenous factors outside of a hospital's control are associated with National Quality Forum–approved ED performance measures, which will have important implications for future benchmarking and public reporting of these data. [Ann Emerg Med. 2012;60:293-298.]

Please see page 294 for the Editor's Capsule Summary of this article.

A **podcast** for this article is available at www.annemergmed.com.

0196-0644/$-see front matter

doi:10.1016/j.annemergmed.2012.01.024

INTRODUCTION

Background

According to the Institute of Medicine, emergency department (ED) crowding is a major public health problem[1] and is associated with patient safety issues, poorer quality, and negative outcomes.[2] The National Quality Forum has endorsed several crowding measures, including median ED length of stay (separated by admitted and discharged patients), waiting times, decision to admit to admission times, and rates of left without being seen.[3] Current plans by the Centers for Medicare & Medicaid Services are to offer higher payments for hospitals who report some of these measures in 2013.[4] Under current specifications, there are no recommended standards for any of the measures and no plan to stratify measures by ED volume or case mix.[3,4] ED performance will likely be measured and compared without consideration of factors out of the control of an ED (ie, exogenous variables).

Importance

Basic assumptions of performance measurement are that hospitals can be compared directly through objective metrics and that the information should be useful to consumers to compare quality and useful to payers to reward better performance. However, this is challenging because differences between patient populations affect the performance measures themselves.[5] Although some differences may be difficult to measure, one solution is risk adjustment, which accounts for observable differences in risk factors (eg, illness severity). Extending this logic to ED throughput, factors outside of a hospital's control may affect performance. For example, a rural ED with 8 beds and 8,000 annual visits is operationally different from an inner-city teaching hospital with 60 beds and

Editor's Capsule Summary

What is already known on this topic

Emergency department (ED) crowding is worsened by hospital and ED efficiency gaps, but the role of external factors is not well described.

What question this study addressed

How do external factors relate to common ED care efficiency metrics?

What this study adds to our knowledge

In a federal sample of US ED visits, median waiting times, ED lengths of stay, and frequency of individuals leaving before care differed widely between geographic locations, case and age mix of patients presenting, and teaching status of the hospital.

How this is relevant to clinical practice

A single set of national ED efficiency benchmarks fails to account for factors that cannot be controlled by individual sites.

120,000 annual visits. Maintaining the same performance (ie, throughput and system responsiveness) at the teaching facility could be more challenging because of system complexity and demand variability. Therefore, one approach to benchmarking is to compare "like" facilities by stratifying data with exogenous variables such as case mix or annual volume.

Goals of This Investigation

We used the National Hospital Ambulatory Medical Care Survey (NHAMCS) to assess the association between exogenous ED-level variables and performance on the National Quality Forum–approved measures. Our goals were to find exogenous ED-level factors to help with benchmarking and public reporting. We sought to create a simple stratification system for ED throughput measures that could be easily understandable by consumers.

MATERIALS AND METHODS

Study Design and Setting

We conducted a study of the 2008 and 2009 NHAMCS to explore facility-level predictors of 4 measures of ED throughput proposed by the National Quality Forum. These include ED length of visit for admitted and discharged patients, waiting time, and rate of left without being seen. The decision to admit to departure from the ED (ie, boarding) was excluded because data on it were not collected in 2008 and greater than 20% of the data were missing in 2009.

Data Collection and Processing

Data came from the 2008 to 2009 NHAMCS, which was conducted by the National Center for Health Statistics. NHAMCS collects data on a nationally representative sample of EDs and ED visits in nonfederal, short-stay general or children's general hospitals, using a multistage probability sample design.[6] The NHAMCS sample has been divided into 16 subsets of approximately equal size. Each subset is contacted during a 4-week reporting period every 16 months (the entire sample does not participate in a given year), and information from approximately 100 patient records is obtained. The sample of EDs used in this study include 334 from the NHAMCS 2009 survey, plus the 98 additional EDs from the 2008 survey that were not also surveyed in 2009. This study was determined by the institutional review board at George Washington University not to be human subjects research.

The primary outcomes were the 4 measures listed above. Because wait time and length of visit within EDs were highly right skewed, we considered their median rather than their mean. To match potential Centers for Medicare & Medicaid Services exclusions, we removed patients with mental health diagnoses and observation admissions from the length of visit for admitted patients.[7] All outcomes were calculated at the ED level. We included several control variables expected to affect ED performance but that might be plausibly exogenous: age mix of ED visits, case mix that consisted of the percentage of ED visits with the 5 most common reasons for visit (digestive, musculoskeletal, cardiovascular, respiratory, and injury), hospital teaching status, annual ED volume, and location in a Metropolitan Statistical Area. We believed that these represented potential variables for reporting stratification because all are exogenous (ie, outside of a hospital's control). These variables are directly available in the NHAMCS, except for teaching status, which was determined from the SMG Hospital Market Database, which serves as the sampling frame for the NHAMCS survey. We assumed that the 8 hospitals with missing teaching status in the SMG file were teaching hospitals if greater than 3% of visits were handled by a resident/intern. Racial composition and insurance status, which have been associated with ED performance in another study, were not included because we thought it would be inappropriate to justify lower performance levels in hospitals with poorer, underinsured populations.[8] Of the 432 EDs in the initial sample, 8 were excluded because they did not have patients with waiting time coded. Of this final sample (424 EDs), 48 did not sample admitted patients, so the sample for length of visit for admitted patients was 376.

Primary Data Analysis

We first identified a set of variables that were plausibly exogenous drivers of ED performance and might serve as a reasonable classification of ED performance. We divided each plausibly exogenous driver of ED performance into quartiles. We then summarized differences in ED performance for each of the exogenous measures, comparing performance among EDs in the upper quartile of the distribution of the measure to the lowest quartile, using Wald χ^2 tests. We considered mean attributes across EDs according to values of the predictor

variables (eg, whether mean performance measures were different for EDs with large visit volumes compared with smaller ones). We consider median wait time within EDs because it is not possible to obtain standard errors for medians across EDs that account for the complex survey design. We then performed multivariate analysis to test the marginal association of each variable with each performance measure, controlling for other exogenous variables. Wait time and visit length across EDs were less right skewed than within EDs. However, because the performance measures were strictly non-negative and still right skewed, properties analysis by ordinary least squares may be biased and inefficient. Accordingly, our multivariate models analyzed performance measures across EDs, using generalized linear models with a γ distribution and log link function. We report goodness of fit measures for generalized linear models.[9] We report results from the generalized linear models as marginal effects, defined as the mean predicted change in the performance measure among EDs in the sample for a 1-unit increase in each independent variable, holding all other independent variables at their observed values. All analyses used sample weights, and standard errors accounted for the complex survey design of NHAMCS with Stata (version 11; StataCorp, College Station, TX).

RESULTS

In 2008 to 2009, median waiting time was 35 minutes (95% confidence interval [CI] 26 to 43 minutes), median length of visit for patients treated but not admitted was 131 minutes (95% CI 121 to 142 minutes), median length of visit for patients admitted was 244 minutes (95% CI 218 to 270 minutes), and rate of left without being seen was 1.3% (95% CI 0.9% to 1.8%) (Table 1). The majority of exogenous variables demonstrated significant bivariate associations with performance measures. For measures of waiting times and overall length of visit, this included ED volume, Metropolitan Statistical Area status, teaching status, and age and case mix. Exogenous variables had a greater proportional effect on waiting time than on length of visit or outcomes of left without being seen. For example, EDs with 10% injury visits had median wait time of 64 minutes, whereas those with more than 20% injury visits had waiting time considerably shorter (26 minutes). Discharged patient length of visit varied from 166 minutes to 113 minutes (32% difference) for the same category, whereas rates of left without being seen were similar. The relationship between the exogenous variables and outcomes were mostly, but not entirely, monotonic. For example, as ED volume increased, lengths of waits and visit increased, though this was not monotonic. There were fewer differences in rates of left without being seen across groups; however, some were significant.

In adjusted analysis, there were several significant predictors for each performance measure (Table 2). The size of the adjusted effect varied between measures and, similar to that of the unadjusted analysis, the directionality was relatively consistent. There were fewer significant predictors of rates of left without being seen compared with other measures. For wait time, the strongest predictor was annual ED visit volume: EDs with greater than 60,000 annual visits had a median wait time 25 minutes longer than that of EDs with less than or equal to 20,000 annual visits. In addition to visit volume and teaching status, 4 other variables, representing EDs in the top quarter according to the percentage of visits among those aged 65 years and over and percentage with reasons for visit as injury or related to musculoskeletal or cardiovascular system, were significantly related to wait time.

Given the large number of significant exogenous variables, we could not construct a simple system for comparing performance measures. Using factors identified as significant would involve reporting performance measures for subsets of EDs with large numbers of permutations of the exogenous variables, so many that this sort of reporting system would be too unwieldy.

LIMITATIONS

We analyzed only 2 years' worth of data. Using different years may yield different results because NHAMCS studies have demonstrated lower performance over time.[10,11] It is also possible that our a priori chosen variables that were available in NHAMCS could have excluded other potentially useful variables. It is also possible that other variables unavailable in NHAMCS (eg, trauma center status, pediatric hospitals) could be helpful in risk stratification. We were unable to study the time from decision to admit to departure (ie, boarding time). Boarding is central to ED crowding; however, the data were not usable because of the high degree of missingness. It also is possible that the definition of some of our calculated variables, such as our definition of teaching hospital at 3% of patients being treated by a resident or intern, overestimated the number of teaching hospitals. In addition, we did not consider interactions or correlations between independent variables in the multivariable model. Finally, some differences in performance across exogenous variables may be statistically but not clinically significant.

DISCUSSION

Several exogenous variables were associated with large differences in performance across 4 ED crowding measures approved by National Quality Forum. Although ED visit volume was the strongest exogenous predictor of performance, several other variables were also important. Therefore, developing a simple, sensible stratification system to compare hospital performance would not adequately characterize the myriad exogenous factors that affect performance. We had initially hypothesized that ED volume would be an adequate adjuster because this is one of the ways that some groups stratify data,[12] but this was not the case. Further complicating matters, because different variables had different effects on the measures (albeit with similar directionality), we can only conclude that a simple way to compare hospitals that accounts for exogenous differences does not exist.

Table 1. Attributes of NHAMCS EDs 2008 to 2009 (N=424 hospitals).*

Variable	Weighted Percentage of Sample	Median Wait Time Among Patients Treated, Minutes	Median Length of Visits Among Patients Treated but Not Admitted, Minutes	Median Length of Visits Among Patients Admitted, Minutes†	Percentage Who Left Without Being Seen
Median (interquartile range)	100	26 (15–40)	118 (98–156)	211 (184–281)	0 (0–1.6)
Mean attributes across EDs (95% percentile)					
All	100	35 (26–43)	131 (121–142)	244 (218–270)	1.3 (0.9–1.8)
Annual ED visit volume					
20,000 or fewer (ref)	50	27 (12–43)	106 (89–124)	212 (166–259)	0.9 (0.1–1.7)
20,001–40,000	24	32 (28–36)	142*** (131–152)	245 (224–266)	1.7 (1.1–2.3)
40,001–60,000	15	52*** (46–59)	166*** (157–175)	298*** (277–320)	2.0** (1.4–2.5)
>60,000	11	49** (41–56)	175*** (163–188)	298*** (267–330)	1.7* (1.3–2.1)
Teaching hospital (ref)	6	33 (24–42)	127 (116–138)	238 (211–265)	1.3 (0.9–1.8)
Nonteaching hospital	94	57*** (46–68)	191*** (173–210)	323*** (284–362)	1.8 (1.2–2.4)
MSA (ref)	64	31 (9–53)	111 (88–135)	188 (162–214)	0.7 (0.2–1.1)
Non-MSA	36	36 (33–40)	142** (135–150)	277*** (243–311)	1.7*** (1.1–2.3)
Percentage of patients aged 18 y and younger					
≤14 (ref)	16	40 (30–51)	154 (131–177)	298 (184–413)	1.2 (0.8–1.6)
14–20	25	43 (14–73)	141 (114–169)	248 (222–274)	1.1 (0.7–1.5)
20–25	24	27** (22–33)	121*** (103–139)	226 (195–257)	0.9 (0.5–1.3)
>25	35	30* (25–35)	121** (112–130)	226 (207–244)	1.9 (0.8–3.0)
Percentage of patients aged 65 y and older					
≤9 (ref)	18	43 (35–51)	145 (140–160)	329 (205–453)	1.6 (1.2–2.1)
9–14	26	47 (20–75)	144 (116–173)	265 (240–289)	2.0 (0.6–3.4)
14–20	24	31** (26–36)	135 (125–145)	224*(207–241)	1.7 (1.1–2.3)
>20	32	22*** (17–27)	111*** (96–126)	206* (175–237)	0.4*** (0.2–0.6)
Percentage of RFV as injury					
≤10 (ref)	16	64 (16–111)	166 (115–218)	309 (194–424)	1.1 (0.8–1.4)
10–15	24	33 (28–38)	141 (133–149)	262 (239–286)	1.6* (1.1–2.1)
15–20	31	29 (25–33)	123 (113–133)	222 (204–240)	1.4 (0.8–1.9)
>20	29	26 (21–31)	113** (100–127)	211* (177–245)	1.2 (0.0–2.4)
Percentage of RFV as digestive system					
≤11 (ref)	27	29 (24–34)	123 (111–135)	272 (189–356)	1.7 (0.4–3.0)
11–14	30	30 (24–36)	119 (106–133)	228 (205–250)	1.0 (0.5–1.5)
14–17	16	31 (26–36)	134 (119–149)	249 (206–293)	1.4 (0.7–2.0)
>17	27	48 (21–75)	152** (125–178)	233 (216–250)	1.4 (1.0–1.8)
Percentage of RFV as musculoskeletal system					
≤10 (ref)	23	49 (16–83)	146 (111–181)	227 (197–258)	0.9 (0.5–1.3)
10–15	32	31 (27–36)	131 (120–142)	234 (207–262)	2.0* (0.9–3.1)
15–20	26	33 (26–39)	133 (122–144)	236 (218–254)	1.5* (1.0–2.0)
>20	19	25 (10–30)	112* (99–126)	290 (183–397)	0.6 (0.3–0.9)
Percentage of RFV as respiratory system					
≤8 (ref)	26	45 (15–75)	148 (117–179)	289 (207–372)	0.8 (0.4–1.2)
8–11	23	28 (20–36)	127 (107–147)	239 (195–283)	1.3 (0.6–1.9)
11–15	25	33 (29–38)	132 (122–143)	242 (222–262)	2.1** (1.0–3.1)
>15	26	29 (24–34)	117** (104–129)	208* (178–237)	1.0 (0.6–1.5)
Percentage of RFV as cardiovascular system					
≤3.5 (ref)	27	44 (16–72)	131 (102–160)	268 (189–346)	0.8 (0.5–1.2)
3.5–5.5	26	33 (27–38)	133 (122–143)	246 (222–270)	1.8 (0.5–3.2)
5.5–7	19	31 (24–38)	131 (116–145)	245 (212–277)	1.5* (0.8–2.2)
>7	28	29 (23–36)	131 (112–151)	222 (188–255)	1.3 (0.8–1.8)

Ref, Reference category; *RFV*, reasons for visit; *MSA*, Metropolitan Statistical Area.
*The symbols *, **, and *** indicate that the difference in this value relative to the ref is significant at the 10%, 5%, and 1% significance level, respectively, according to a Wald test.
†The number of hospitals for admitted patients is 376.

The logical question is, how should the public reporting of these data be conducted when they are eventually presented to the public? One possibility is to just report unadjusted data. However, if consumers use unadjusted information to inform hospital choice, they could choose smaller, rural hospitals because ED performance is better. Although this may be appropriate, we think that consumers should have an explanation of factors affecting wait time, as well as other factors such as quality of care ratings, which may give a fuller picture to properly inform ED choice.

Table 2. Predictors from generalized linear model regression among patients treated in NHAMCS EDs, 2008 to 2009.*

	95% CI			
Variable	**Median Wait Time Among Patients Treated, Minutes**	**Median Length of Visits Among Patients Treated but Not Admitted, Minutes**	**Median Length of Visits Among Patients Admitted, Minutes†**	**Percentage Who Left Without Being Seen**
Annual ED visit volume (relative to 20,000 or fewer)				
20,001–40,000	8* (−1 to 17)	33*** (11 to 55)	16 (−37 to 70)	0.6 (−0.4 to 1.6)
40,000–60,000	34*** (21 to 47)	59*** (38 to 80)	60* (−4 to 124)	1.0 (−0.6 to 2.5)
>60,000	25*** (13 to 38)	63.8*** (40 to 88)	54 (−12 to 119)	0.5 (−0.7 to 1.7)
Teaching hospital (relative to nonteaching hospital)	8** (1 to 15)	20*** (7 to 34)	36.1** (1 to 71.2)	0.1 (−0.3 to 0.6)
Non-MSA (relative to MSA)	19 (−5 to 27)	7 (−15 to 29)	−47* (−98 to 3)	−0.3 (−1.2 to 0.6)
Top quartile (>25%) of % of patients younger than 18 y (relative to bottom 3 quartiles)	−6 (−15 to 3)	−14.3* (−31 to 2)	−7 (−34 to 20)	0.4 (−0.2 to 1.0)
Top quartile (>20%) of % of patients aged 65 y or older (relative to bottom 3 quartiles)	−17*** (−23 to –4)	−17* (−37 to 3)	−15 (−48 to 18)	−1.0*** (−1.5 to −0.6)
Top quartile (>20%) of % of RFV as injury (relative to bottom 3 quartiles)	−10*** (−17 to −4)	−26*** (−40 to −11)	−42** (−83 to −1)	−0.4 (−1.0 to 0.2)
Top quartile (>17%) of % of RFV as digestive system (relative to bottom 3 quartiles)	10 (−3 to 22)	19** (1 to 38)	−9 (−39 to 21)	0.1 (−0.5 to 0.7)
Top quartile (>20%) of % of RFV as musculoskeletal system (relative to bottom 3 quartiles)	−9** (−16 to −1)	−18** (−34 to −1)	50 (−14 to 115)	−0.4 (−1.0 to 0.2)
Top quartile (>15%) of % of RFV as respiratory system (relative to bottom 3 quartiles)	−3 (−8 to 3)	−18*** (−28 to −9)	−44*** (−75 to −12)	−0.6** (−1.1 to −0.1)
Top quartile (>7%) of % of RFV as cardiovascular system (relative to bottom 3 quartiles)	−7* (−14.8 to 0.6)	−10 (−25 to 6)	−30* (−65 to 5)	−0.0 (−0.6 to 0.5)
Pseudo R^2	0.077	0.191	0.098	0.039

*The table reports marginal effects from generalized linear model regression with γ distribution assumed. The symbols *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample size is 424 hospitals.
†The number of hospitals for admitted patients is 376.

A separate issue is how these measures may be used to show "high performers." A uniform approach to judging flow across all hospitals was implemented by the United Kingdom in 2004 by requiring all hospitals to reduce length of stay to be under 4 hours; however, recently this standard has been relaxed.[13] Our findings demonstrate that this may unfairly identify high- or low-performing hospitals because of factors outside their control. Also, given reported relationships between racial and insurance composition and performance, it could widen disparities because EDs serving the poor could receive lower payments if these measures are used in any future pay-for-performance incentives.[8] One solution might be to create a more complex system to adjust for exogenous factors that decreases disparities. On the other hand, a limitation of allowing different benchmarks for certain hospitals is that it could potentially excuse longer waiting times in certain hospitals, decreasing the incentive to improve ED flow.

In conclusion, several exogenous factors were associated with ED crowding measures, which will have important policy implications for public reporting and any future pay-for-performance incentives.

Supervising editor: Donald M. Yealy, MD

Author contributions: JMP and SLD conceived this study and drafted the article. SLD and TH conducted the statistical analysis. TH provided critical revisions to the article. JMP takes responsibility for the paper as a whole.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other

relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Publication dates: Received for publication October 25, 2011. Revision received January 18, 2012. Accepted for publication January 25, 2012. Available online May 23, 2012.

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Address for correspondence: Jesse M. Pines, MD, MBA, MSCE, E-mail jesse.pines@gmail.com.

REFERENCES

1. Institute of Medicine. The future of emergency care in the United States health system. *Acad Emerg Med*. 2006;13:1081-1085.
2. Bernstein SL, Aronsky D, Duseja R, et al. The effect of emergency department crowding on clinically oriented outcomes. *Acad Emerg Med*. 2009;16:1-10.
3. National Health Quality Forum. National Quality Forum Endorses Measures to Address Care Coordination and Efficiency in Hospital Emergency Departments. Available at: http://www.qualityforum.org/News_And_Resources/Press_Releases/2008/National_Quality_Forum__Endorses_Measures_to_Address_Care_Coordination_and_Efficiency_in_Hospital_Emergency_Departments.aspx. Accessed January 16, 2012.
4. Centers for Medicare & Medicaid Services. CMS-1504-FC hospital outpatient prospective payment system and CY 2011 payment rates. November 2, 2010. Hospital outpatient regulations and notices. Available at: http://www.cms.gov/hospitaloutpatientpps/hord/itemdetail.asp?itemid=CMS1240960. Accessed January 16, 2012.
5. Iezzoni LI. The risks of risk adjustment. *JAMA*. 1997;278:1600-1607.
6. McCaig LF, McLemore T. Plan and operation of the National Hospital Ambulatory Medical Survey. Series 1: programs and collection procedures. *Vital Health Stat 1*. 1994;34:1-78.
7. Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/QualityMeasures/Downloads/EH_EDThroughputStratificationTable.pdf. Accessed January 16, 2012.
8. Pines JM, Russell AR, Hollander JE. Racial disparities in emergency department length of stay for admitted patients in the United States. *Acad Emerg Med*. 2009;16:403-410.
9. Zheng B, Agresti A. Summarizing the predictive power of a generalized linear model. *Stat Med*. 2000;19:1771-1781.
10. Wilper AP, Woolhandler S, Lasser KE, et al. Waits to see an emergency department physician: US trends and predictors, 1997-2004. *Health Aff (Millwood)*. 2008;27:w84-95.
11. Herring A, Wilper A, Himmelstein DU, et al. Increasing length of stay among adult visits to US emergency departments, 2001-2005. *Acad Emerg Med*. 2009;16:609-616.
12. Available at: http://edbenchmarking.org/. Accessed January 16, 2012.
13. Mason S. United Kingdom experiences in evaluating performance and quality in emergency medicine. *Acad Emerg Med*. 2011;12:1234-1238.